# Multi-scale keypoint annotation - a biological approach

Miguel Farrajota[1,2], João Rodrigues[1,2] and J.M.H. du Buf[1]
[1]Vision Laboratory, Institute for Systems and Robotics (ISR), University of the Algarve, Faro, Portugal
[2]Instituto Superior de Engenharia, University of the Algarve, Faro, Portugal
elsio_farrajota@hotmail.com, {jrodrig, dubuf}@ualg.pt

## Abstract

*The primary visual cortex employs simple, complex and end-stopped cells to create a scale space of 1D singularities (lines and edges) and of 2D singularities (line and edge junctions and crossings called keypoints). In this paper we show first results of a biological model which attributes information of the local image structure to keypoints at all scales, ie junction type (L, T, +) and main line/edge orientations. Keypoint annotation in combination with coarse to fine scale processing facilitates various processes, such as image matching (stereo and optical flow), object segregation and object tracking.*

## 1. Introduction

Our own framework of the visual cortex is based on simple, complex and end-stopped cells in V1. Simple and complex cells serve to detect lines and edges, whereas end-stopped cells yield keypoints at line and edge junctions and crossings. The two scale spaces (LE and KP) can be combined for invariant object recognition [5]. However, object recognition by dynamically routing features of unknown input objects to features of known objects in memory implies that the visual cortex is blocked until the routing networks are released; hence, only one object can be tested at any time. This sequential process, driven by Focus-of-Attention, is bootstrapped by fast gist vision: global scene gist and local object gist, the latter for creating a first layout map of likely object categories at approximate locations in the visual field. Local gist and layout are related to object segregation and recognition, a typical chicken-or-egg problem: for recognition the object must be segregated but segregation is impossible without knowing what the object is. Local gist can be based on color and texture features, but disparity (stereo; depth) and optical flow are very important to bind together different object parts.

Our goal is to develop a biological model which integrates gist vision with scene and object recognition. These problems are also tackled in computer vision, applying many different approaches. Here we focus on keypoints and their annotation for solving the correspondence problem. Instead of working with mere point clouds in combination with mathematical methods [1], there is growing interest in using additional information, e.g. [2, 3, 6]. There are three reasons for doing this: (1) By definition, keypoints at any scale are caused by junctions of lines and edges at the same scale, and this line/edge information is tractable for keypoint annotation. (2) At coarser scales, the larger receptive field sizes of the cells will lead to more interference effects and therefore less lines and edges which will be detected [5], but the same effects will occur in image pairs (stereo for disparity; sequential for optical flow). (3) Going from a very coarse scale to the finest one, one keypoint is located at the center of an object, then more keypoints are found at object parts, and finally most keypoints are at object (and contour) details. (4) Matching annotated keypoints is much easier than matching mere point clouds, especially when one starts at a coarse scale and the matching results are used to steer the matching at a finer scale, the process being applied repeatedly until the finest scale.

## 2. Keypoint annotation

We apply keypoint detection with end-stopped cells and sophisticated inhibition schemes as detailed in [4], which yields keypoint positions in floating-point coordinates (superresolution). Here we use simple, complex and end-stopped cells at eight scales and with eight orientations. Receptive fields of simple cells are modeled by complex Gabor functions with sine and cosine components, complex cells by the modulus. The scale is given by $\lambda$, the wavelength of the simple cells in pixels: $\lambda = 6, 12, 18, \ldots 48$.

For annotating any detected keypoint, responses of simple and complex cells are analysed at three distances from its position, over orientation intervals around 8 main orientations. If $\Delta\phi = 2\pi/8$, main orientations are $\phi_k = k\Delta\phi$ with $k = 0, \ldots, 7$. With $\delta\phi = \Delta\phi/2$, the 8 orientation intervals are $\Phi_k = \phi_k \pm \delta\phi$. The three distances are $\lambda/2$, $\lambda$

and $2\lambda$.

For each $\Phi_k$, the maximum response $R_k$ within the interval and at the three distances is detected, and their maximum $\hat{R} = \max_k R_k$ and the average $\bar{R} = 1/8 \sum_k R_k$ are computed. Then a dynamic thresholding is applied: (1) If $\bar{R} < 0.5\hat{R}$, any orientation $k$ for which $R_k > 0.75\hat{R}$ will be used for keypoint annotation. (2) If $0.5\hat{R} < \bar{R} < 0.8\hat{R}$, any $k$ for which $R_k > 0.9\hat{R}$ will be used. (3) If $\bar{R} > 0.8\hat{R}$, any $k$ for which $R_k > 0.98\hat{R}$ will be used. The above values were determined by analysing simple objects like triangles, squares and polygons. Resulting orientations $k_i$ are attributed to the keypoint, plus the junction type (L, T, +). There is one exception: isolated points and blobs, especially at very coarse scales, are also detected but they are not caused by line/edge junctions. Such keypoints are labeled "blob" without orientations.

Figure 1 illustrates the keypoint annotation process. It shows three scales, $\lambda = 6$ (row 1 to 3), $\lambda = 12$ (row 4 to 6) and $\lambda = 24$ (row 7 to 9), with at each scale three orientations of all 8. From the left to the right it shows responses of even and odd simple cells plus complex cells. In the case of simple cells, dark values are negative and bright ones are positive. Responses of complex cells are always positive. Also shown are detected keypoints with (in red) the three distances $\lambda/2$, $\lambda$ and $2\lambda$ at which the responses are tested. Responses and keypoints shown are at and close to the upper-left corner of the bottom object in Fig. 2 (top-left).

Figure 2 presents first multiscale keypoint annotation results. The left column shows, top to bottom, frames 1, 5, 10, 15 and 20 of a test sequence with three moving objects, the latter four in combination with the first frame to show the movements. The second column shows keypoints detected in frame 1 at $\lambda = 6$, 12, 24, 30 and 42. The last column shows the same keypoints with annotated orientations. (Note: the pdf viewer should be used with a large magnification in order to see all details.)

## 3. Tracking and segregation

In stereo and optical flow, annotated keypoints in left-right and successive frames are compared in an area with radius $\lambda$ according to the keypoints' scales, starting at $\lambda = 48$ and going to $\lambda = 6$, and the corresponding vectors are computed. In object segregation and object tracking the processing is the same, matching keypoints being linked over and within scales by using a multi-branch tree structure (biological implementation to be developed).

Figure 3 shows the results of object tracking, at top-left from frame 1 to 10 and at top-right from frame 1 to 20. The bottom image shows the movements of keypoints of the segregated objects coded with different colors.

Being first results, they are not yet perfect but most

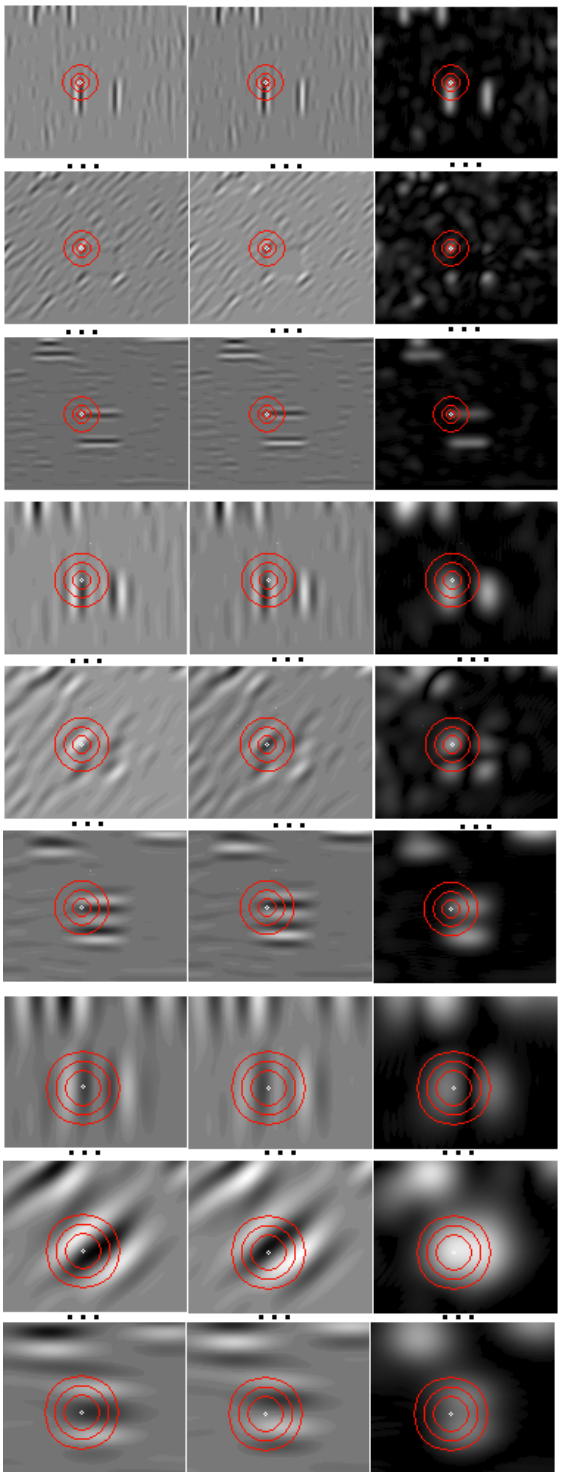

**Figure 1. Top to bottom: 3 scales, with 3 orientations per scale. Left to right: even and odd simple and complex cells, plus detected keypoints with (in red) the 3 distances at which the responses are tested. See text.**

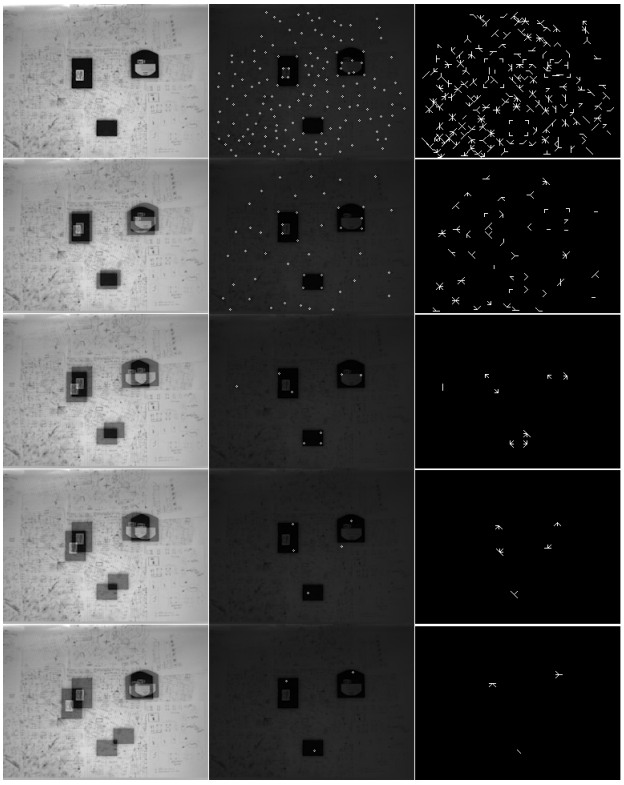

**Figure 2. Left: five frames of a test sequence with three moving objects. Middle: keypoints detected in the first (top) frame at five scales. Right: the keypoints with annotated orientations. See text.**

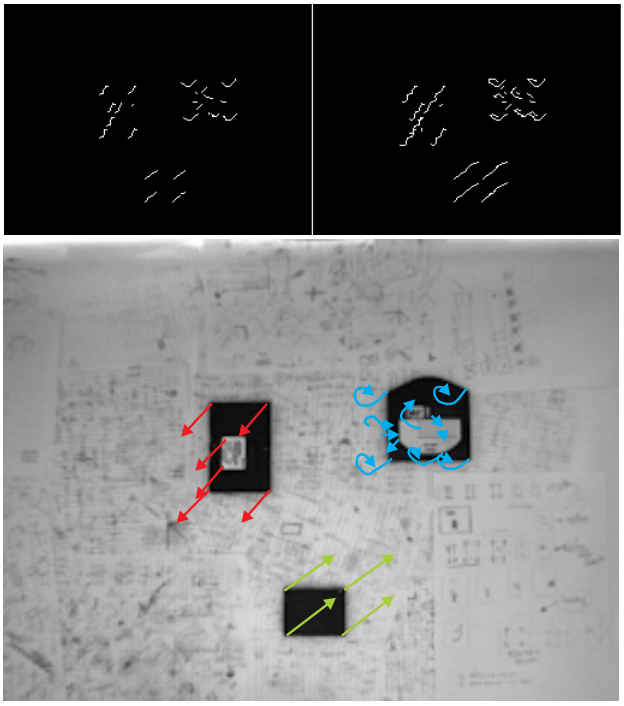

**Figure 3. Results of object tracking, frames 1 to 10 (top-left) and frames 1 to 20 (top-right). Bottom: keypoint movements of segregated objects coded with different colors.**

keypoints and orientations are correct and ongoing research concentrates on further optimizations. In addition to the tracking of keypoints we can obtain the segregation of the objects through the objects' keypoints within the same motion region, in Fig. 3 (bottom) indicated by different colors. This segregation process is now being combined with the segregation information obtained from multi-scale keypoints without motion tracking [4] and from the multi-scale line-edge representation [5].

***Acknowledgements:*** This research is partly supported by the FCT (ISR/IST pluri-annual funding) through the POS Conhecimento Program, which includes FEDER funds, and by the FCT project PTDC/EIA/73633/2006 - SmartVision: active vision for the blind.

## References

[1] P. Dorst at al. Optic flow using multi-scale anchor points. *Proc. 13th Int. Conf. Computer Analysis of Images and Patterns, Mnster, Germany, September 2-4*, LNCS 5702:1104–1112, 2009.

[2] V. Garcia et al. Region-of-interest tracking based on keypoint trajectories on a group of pictures. *Proc. Int. Workshop on Content-Based Multimedia Indexing, Bordeaux, France, 25-27 June*, pages 198–203, 2007.

[3] J. Mooser et al. Fast simultaneous tracking and recognition using incremental keypoint matching. *Proc. 4th Int. Symposium on 3D Processing, Visualization and Transmission, Georgia Institute of Technology, Atlanta, GA, USA, June 18 - 20*, page 8, 2008.

[4] J. Rodrigues and J. du Buf. Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. *BioSystems*, 86:75–90, 2006.

[5] J. Rodrigues and J. du Buf. A cortical framework for invariant object categorization and recognition. *Cognitive Processing*, 10(3):243–261, 2009.

[6] R. Trichet and B. Merialdo. Keypoints labeling for backgroung substraction in tracking applications. *Proc. IEEE Int. Conf. on Multimedia and Expo, Hannover, Germany, June 23-26*, pages 1301–1304, 2008.